Take-home Final / Problem Set

Instructions:

1. You are to complete this assignment on your own, without discussing its content or your results with anyone else.
2. Please turn in clear and concise solutions to the questions asked, in addition to a table of output (e.g., posterior means and standard deviations).
3. In addition to the above, please include a printout of your code.
4. This assignment is due on Monday, October 17, by 9:00 am. Please DO NOT turn this into my mailbox, but instead, deliver it to me personally. You are welcome to submit your answers and results early - if you choose to do so, please drop them off in person to my office (RAWLS 4076).

Thanks and good luck!

As discussed in class, the Poisson distribution, used for modeling count outcomes, has a density function given by:

$$p(y|\theta) = \frac{\theta^y \exp(-\theta)}{y!}, \quad y = 0, 1, 2, \ldots$$

Consider use of a *two-component mixture of Poissons model*, similar to that described in your lecture notes for Gaussian (normal) mixtures. Let the probability associated with the first component be given as π, and let θ_1 and θ_2 denote the parameters of each component of the Poisson mixture.

(**a: 10 Points**) After augmenting the parameter space with a set of component indicator variables $\mathbf{z}$, write down the conditional (on $\mathbf{z}$ and the parameters θ_1 and θ_2) likelihood function for this model.

(**b: 10 Points**) Describe reasonable (and computationally convenient) priors to use for θ_1, θ_2 and π.

(**c: 10 Points**) Write down the (augmented) joint posterior distribution for this model.

(**d: 20 Points**) Outline a Gibbs sampler for fitting this two component Poisson mixture. Describe each conditional posterior distribution in full detail.

(**e: 50 Points**) Data on the course website, bikes.txt, describes the number of instances of bike traffic on a fixed roadway segment during the course of one month. Using your results in part (d) above, write a program to fit a two-component Poisson mixture model to this data. Your code should be used to:

- Calculate posterior means and posterior standard deviations associated with all parameters of this model. Comment on how long your posterior simulator was run, and your choice of burn-in period.
- Plot a histogram of the observed bike traffic data, along with an estimate of that density from your model (to make things simple, you can just evaluate the density at parameter posterior means).